Integrated Access to Heliospheric and Magnetospheric Data

Jan Merka[1] A. Szabo[2] T. W. Narock[1]

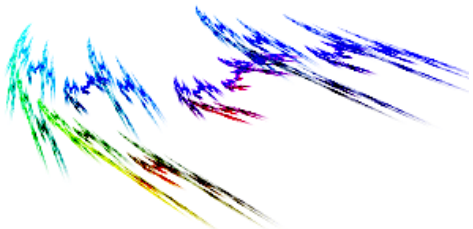
[1]GEST, University of Maryland Baltimore County, Baltimore, MD (jan.merka@gsfc.nasa.gov)

[2]NASA Goddard Space Flight Center, Greenbelt, MD

AGU Joint Assembly, Acapulco, Mexico

The Need for Integrated Data Access

- Hundreds of data products exist (VSPO currently shows 387)
- Knowing that a data source exists and where is often not enough to employ the data for research
- As a solution, NASA Heliophysics Division has established a group of virtual observatories (VOs)
- Virtual Heliospheric and Magnetospheric Observatories (VHO and VMO) aim to provide integrated access to heliospheric and magnetospheric data sets.
- VHO and VMO share the same architecture and provide discipline-tuned interfaces



Principal VxO Features

Definition

A VO is a service that unites services and/or multiple data providers as a suite of a software applications on a set of computers that allows users to uniformly find, access, and use resources (data, software, document, and image products and services using these) from a collection of distributed product and service providers.

Key features:

1. Distributed environment
2. Modular structure
3. Uniform interface and treatment of resources (data, services, etc.)
4. Standard data descriptions
5. Standard communication among services and VOs

Convenience is Important for Success

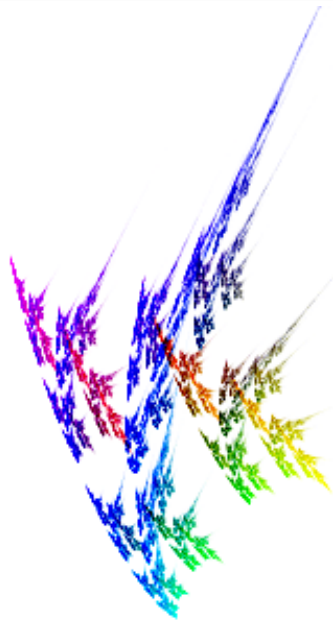
Example

Google succeeded in web-search partly by avoiding the exact search language of AltaVista in favor of a tool which was fast, easy to use, and displayed most of the right results in mostly the right order.

Bill Softky, The Register

VHO/VMO aim to offer

- User convenience
  - Intuitive web interface
  - Unified information presentation and retrieval
  - Fast response time or warning before time-demanding actions
  - Data value queries
- Accurate results
  - Accuracy is limited by metadata availability
- Integrate with services and other VxOs



Convenience is Important for Success

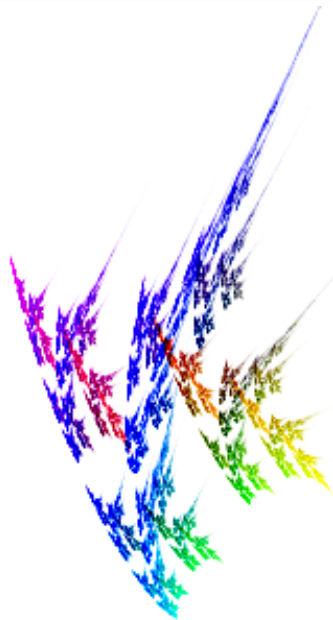
Example

Google succeeded in web-search partly by avoiding the exact search language of AltaVista in favor of a tool which was fast, easy to use, and displayed most of the right results in mostly the right order.

Bill Softky, The Register

VHO/VMO aim to offer

- User convenience
  - Intuitive web interface
  - Unified information presentation and retrieval
  - Fast response time or warning before time-demanding actions
  - Data value queries
- Accurate results
  - Accuracy is limited by metadata availability
- Integrate with services and other VxOs



Convenience is Important for Success

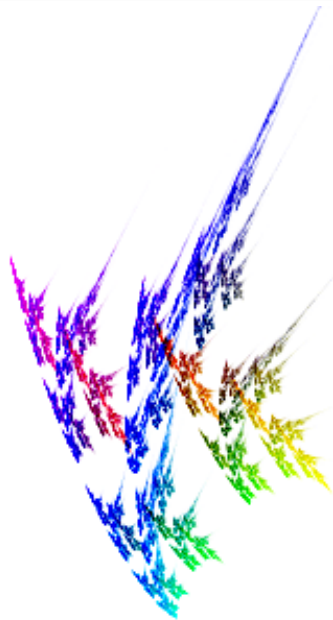
Example

Google succeeded in web-search partly by avoiding the exact search language of AltaVista in favor of a tool which was fast, easy to use, and displayed most of the right results in mostly the right order.

Bill Softky, The Register

VHO/VMO aim to offer

- User convenience
  - Intuitive web interface
  - Unified information presentation and retrieval
  - Fast response time or warning before time-demanding actions
  - Data value queries
- Accurate results
  - Accuracy is limited by metadata availability
- Integrate with services and other VxOs

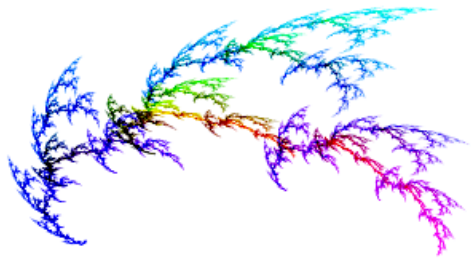


Query Types Supported by SPASE Metadata

- SPASE data model allows for detailed description of resources from observatory to data product level
  - Search for measurement type, instrument type, observed regions, text patterns, ...
  - This information is insufficient for finding specific data granules
- SPASE also describes time extent of data files (granules)
  - Find granules matching specified time interval(s)

But we want more!

- Search for specific positions or regions
- Find data based on the observed parameter values

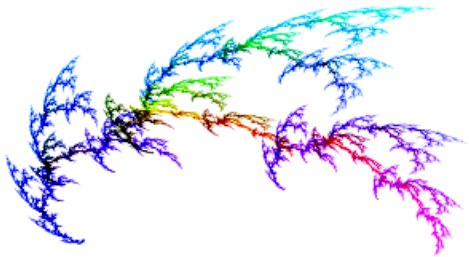


Query Types Supported by SPASE Metadata

- SPASE data model allows for detailed description of resources from observatory to data product level
  - Search for measurement type, instrument type, observed regions, text patterns, ...
  - This information is insufficient for finding specific data granules
- SPASE also describes time extent of data files (granules)
  - Find granules matching specified time interval(s)

But we want more!

- Search for specific positions or regions
- Find data based on the observed parameter values

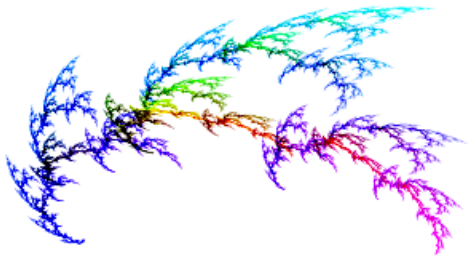


Query Types Supported by SPASE Metadata

- SPASE data model allows for detailed description of resources from observatory to data product level
  - Search for measurement type, instrument type, observed regions, text patterns, ...
  - This information is insufficient for finding specific data granules
- SPASE also describes time extent of data files (granules)
  - Find granules matching specified time interval(s)

But we want more!

- Search for specific positions or regions
- Find data based on the observed parameter values

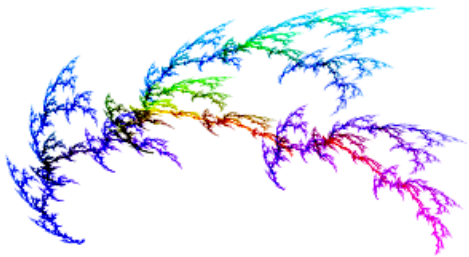


Query Types Supported by SPASE Metadata

- SPASE data model allows for detailed description of resources from observatory to data product level
  - Search for measurement type, instrument type, observed regions, text patterns, ...
  - This information is insufficient for finding specific data granules
- SPASE also describes time extent of data files (granules)
  - Find granules matching specified time interval(s)

But we want more!

- Search for specific positions or regions
- Find data based on the observed parameter values



Extensions to SPASE Model Enable More Advanced Queries

- Employ the *Extension* element of SPASE data model to combine SPASE with VOTable descriptions
- VOTable effectively captures tabular information in XML format
- VHO/VMO uses the *Extension* to obtain more information about
  - Ephemeris data products
    - ★ Provide location of the observation at variable time step
  - Other data products (observed parameters)
    - ★ Provide statistical summary of parameter values over variable-length time intervals
    - ★ The statistical parameters are, for example, average, median, standard deviation, minimun, maximum values, and data availability ratio.

Newly enabled queries

- Find thermal plasma measurements where $R_{GSE} < 10$ R_E and $X_{GSE} > 5$ R_E.
- Find time intervals when (average) IMF $B_Z < 0$ nT.

Extensions to SPASE Model Enable More Advanced Queries

- Employ the *Extension* element of SPASE data model to combine SPASE with VOTable descriptions
- VOTable effectively captures tabular information in XML format
- VHO/VMO uses the *Extension* to obtain more information about
  - Ephemeris data products
    - ★ Provide location of the observation at variable time step
  - Other data products (observed parameters)
    - ★ Provide statistical summary of parameter values over variable-length time intervals
    - ★ The statistical parameters are, for example, average, median, standard deviation, minimun, maximum values, and data availability ratio.

Newly enabled queries

- Find thermal plasma measurements where $R_{GSE} < 10$ R_E and $X_{GSE} > 5$ R_E.
- Find time intervals when (average) IMF $B_Z < 0$ nT.

Prototype VHO/VMO Web Interface

Query Builder Version: 0.6

- Search criteria
  - Measurement type
    - Ephemeris
      - [?] AMPTE Magnetic Field Experiment (MFE) and Positions
      - [?] WIND Ephemeris
        - [?] WIND Definitive Ephemeris
        - [?] WIND Predicted Ephemeris
    - Magnetic field
      - [?] AMPTE Magnetic Field Experiment (MFE) and Positions
      - [?] WIND Magnetic Field Investigation (MFI)
  - Restrictions
    - Date/Time
    - Position
      - GSE
      - GSM
      - SM

Submit **Current Query**

[-] Magnetic field: AMPTE Magnetic Field Experiment (MFE) and Positions

[-] [+] Position in GSE:

	X [R_E]	Y [R_E]	Z [R_E]
Minimum (>=)	0	0	-INF
Maximum (<=)	5	5	INF

[-][+] Position in GSE:

	X [R_E]	Y [R_E]	Z [R_E]
Minimum (>=)	-5	-5	-INF
Maximum (<=)	0	0	INF

Organized Results

[Back to Query Builder]

Search Results

Found 131 data files.

Displaying 50 records (1—50).

Next »

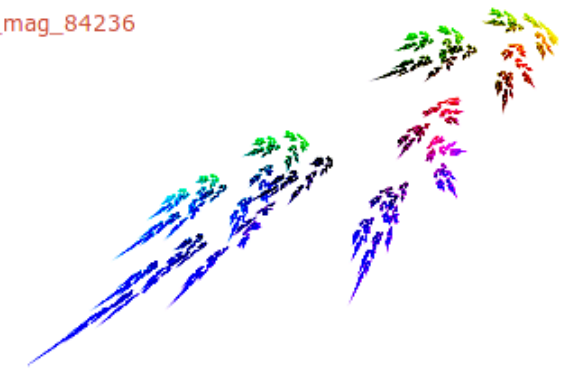
	UTC Time Interval	Data File	Info	Type
1	1984-08-21 03:15:48 — 1984-08-21 04:01:00 1984-08-21 05:31:23 — 1984-08-21 05:31:23 1984-08-21 07:31:51 — 1984-08-21 07:31:51 1984-08-21 19:05:00 — 1984-08-21 19:20:04 1984-08-21 19:35:09 — 1984-08-21 19:35:09 1984-08-21 21:05:33 — 1984-08-21 21:20:37	sa_mag_84234.dat	G D I O	Ephemeris Magnetic field
2	1984-08-22 10:47:46 — 1984-08-22 11:17:54 1984-08-22 12:48:19 — 1984-08-22 12:48:19	sa_mag_84235.dat	G D I O	Ephemeris Magnetic field
3	1984-08-23 02:15:28 — 1984-08-23 02:45:36 1984-08-23 03:00:41 — 1984-08-23 03:00:41 1984-08-23 04:31:05 — 1984-08-23 04:31:05 1984-08-23 18:04:29 — 1984-08-23 18:04:29 1984-08-23 18:19:34 — 1984-08-23 18:19:34 1984-08-23 18:34:39 — 1984-08-23 18:34:39 1984-08-23 20:05:03 — 1984-08-23 20:20:07	sa_mag_84236.dat	G D I O	Ephemeris Magnetic field
4	1984-08-24 09:33:39 — 1984-08-24 10:03:47 1984-08-24 10:18:52 — 1984-08-24 10:18:52 1984-08-24 11:49:16 — 1984-08-24 11:49:16	sa_mag_84237.dat	G D I O	Ephemeris Magnetic field
5	1984-08-25 01:15:08 — 1984-08-25 01:30:12 1984-08-25 01:45:17 — 1984-08-25 01:45:17 1984-08-25 02:00:21 — 1984-08-25 02:00:21 1984-08-25 03:30:42 — 1984-08-25 03:30:42 1984-08-25 16:49:06 — 1984-08-25 17:49:22 1984-08-25 19:04:43 — 1984-08-25 19:04:43	sa_mag_84238.dat	G D I O	Ephemeris Magnetic field
6	1984-08-26 08:31:45 — 1984-08-26 08:31:45 1984-08-26 08:46:49 — 1984-08-26 09:16:57	sa_mag_84239.dat	G D I O	Ephemeris

Metadata Display

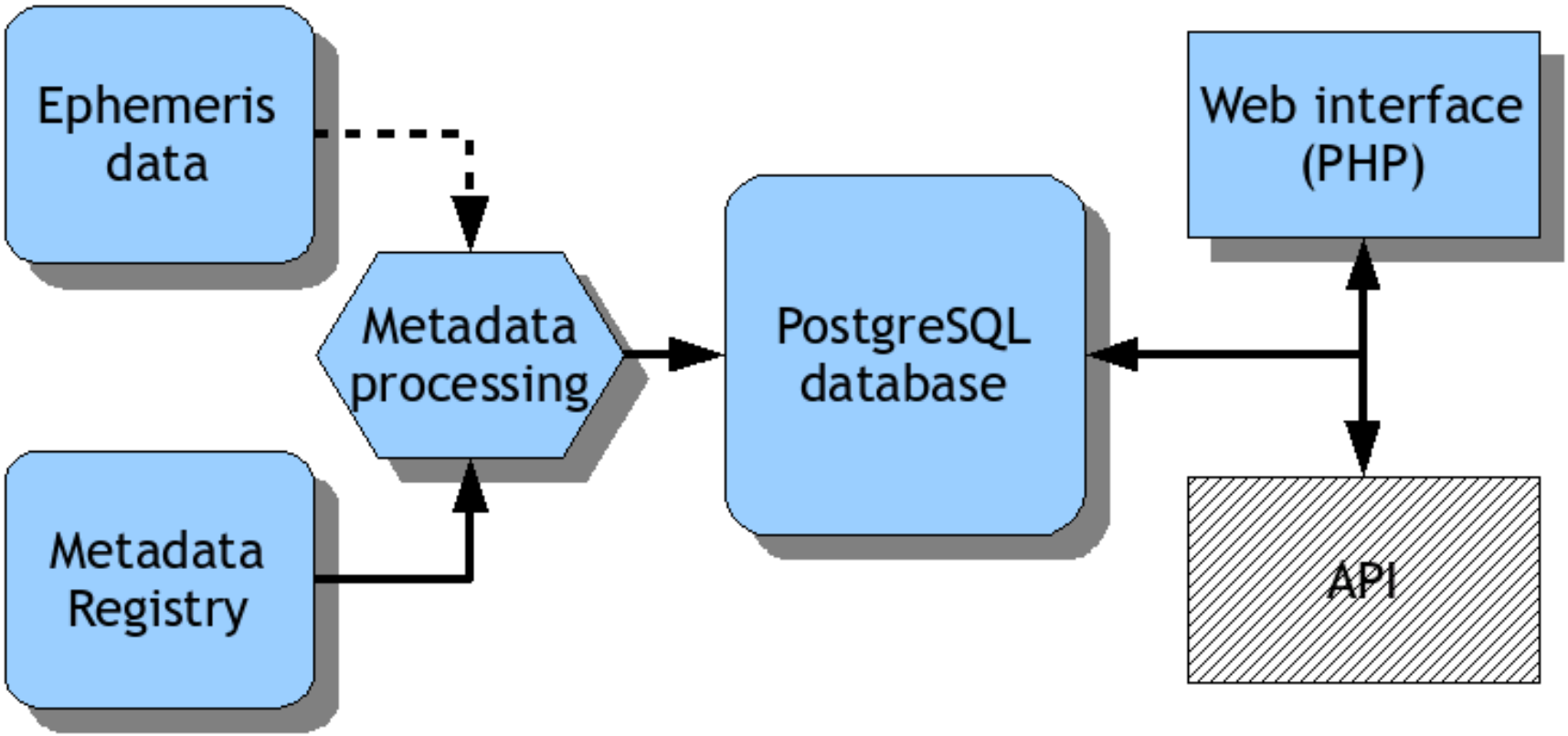
- Detailed information about resource is stored in XML format
- VHO/VMO extracts such information and displays it conveniently formated for viewing (no XML tags, . . .)
- Metadata about a data granule (file) from AMPTE-CCE spacecraft:

Data File Details [Granule Resource]

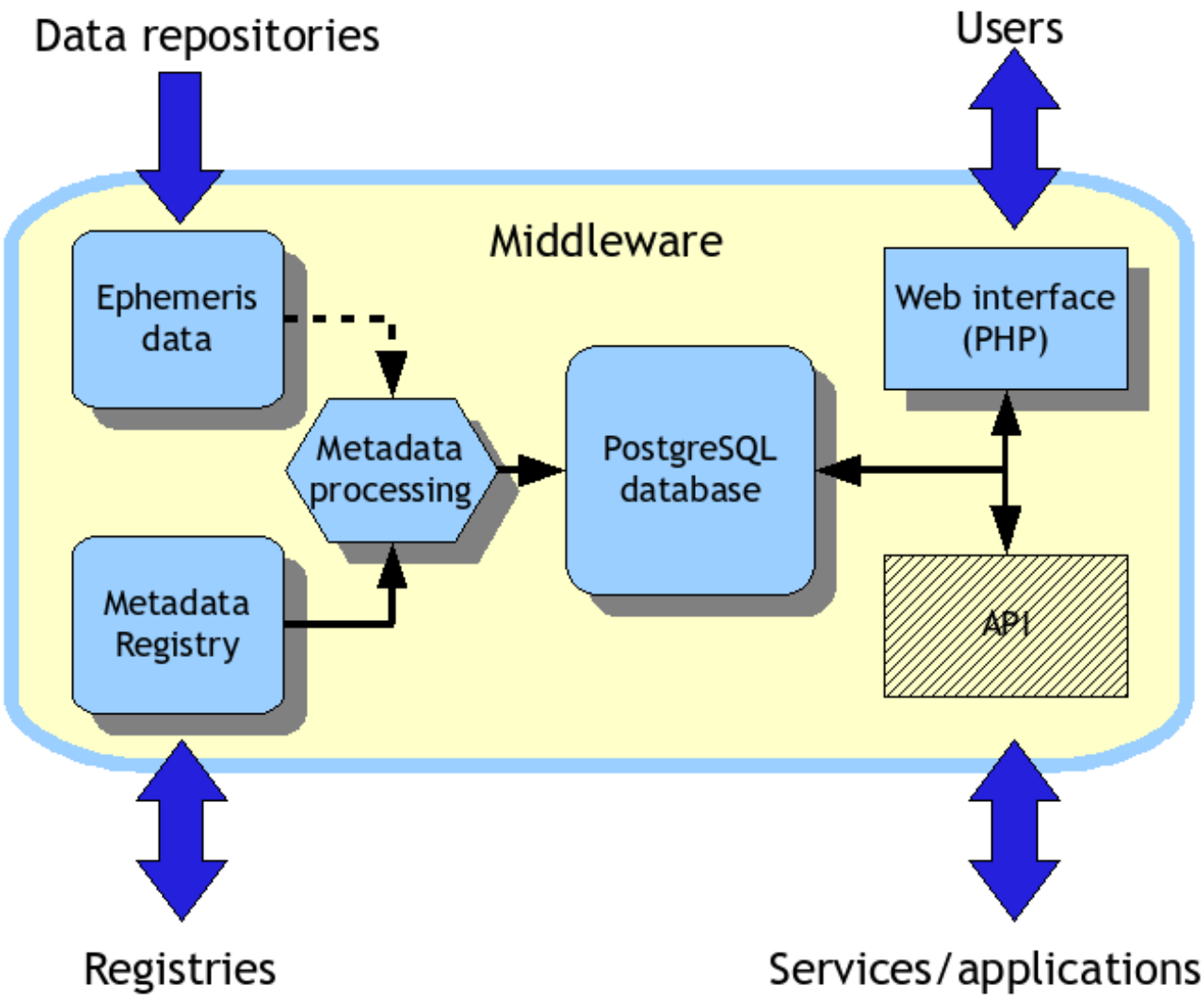
Resource ID	spase://GSFC/obs/AMPTE_CCE/MFE/AMPTE_CCE_MAG/sa_mag_84236
Parent ID	spase://GSFC/obs/AMPTE_CCE/MFE/AMPTE_CCE_MAG
Release date	2007-05-15 12:48:05
Start date	1984-08-23 00:00:00
Stop date	1984-08-23 23:59:59
URL	sa_mag_84236.dat
File size	1125239 Bytes
MD5 checksum	50a762b0454e47a8492158d041dab05c



Middleware Architecture



Middleware Connectivity



Virtual Observatory Query Language (VOQL)

How will services communicate with VxOs?

Future services will have to send queries to VxOs or other services and then retrieve query results. The VxOs already use standard set of terms (SPASE) for data description but still need to define how to construct queries and result sets using these terms.

VO Query Language:

- A small piece of software that serves as a front-end to VOs, services and data centers.
- It could serve as a common standardized interface language
- Each service would require only a single modification to connect to the pool of VOQL-speaking VxO environment.

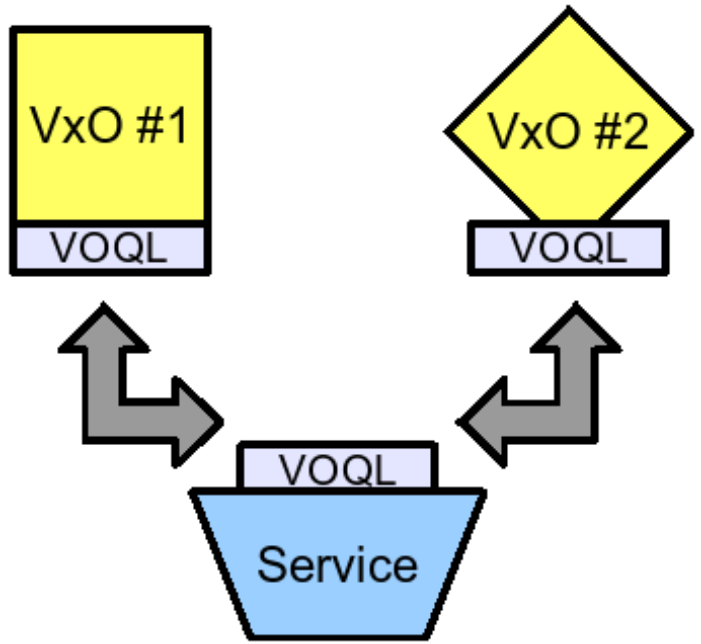
Virtual Observatory Query Language (VOQL)

How will services communicate with VxOs?

Future services will have to send queries to VxOs or other services and then retrieve query results. The VxOs already use standard set of terms (SPASE) for data description but still need to define how to construct queries and result sets using these terms.

VO Query Language:

- A small piece of software that serves as a front-end to VOs, services and data centers.
- It could serve as a common standardized interface language
- Each service would require only a single modification to connect to the pool of VOQL-speaking VxO environment.



VOQL Example

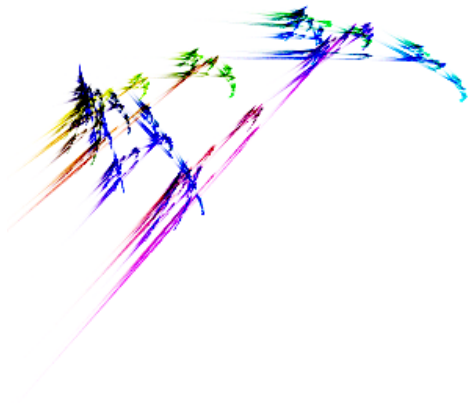
The VOQL is developed in collaboration with other VxO groups and service providers.

Query

(Observed Region = near-Earth heliosphere) AND (magnetic field $B_Z^{GSE} < 0$ nT)

```
<VOQL>
   <Qterm>
        <ID>0</ID>
        <SPASE>Heliosphere.NearEarth</SPASE>
   </Qterm>
   <Qterm>
        <ID>1</ID>
        <SPASE>MagneticField</SPASE>
        <SPASE>Z</SPASE>
        <SPASE>GSE</SPASE>
        <QueryDataUnits>nT</QueryDataUnits>
        <LT>0</LT>
   </Qterm>
</VOQL>
```

VHO/VMO Integration



- Common architecture and components
  - Faster development and more thorough testing
  - Differences between VHO and VMO
    - ★ Discipline-specific interface modifications
    - ★ Default query constraints (magnetosphere $\times$ heliosphere)
    - ★ Coordinate systems selection (GSE, SM, GSM $\times$ GSE, RTN, HGI)
    - ★ VMO is unique with two centers of activity at NASA/GSFC and UCLA working on complementary tasks
  - Some components (e.g. registry from VMO/U) are developed and shared with other V$\times$O groups
- Provide solar wind/IMF data relevant to VMO queries
  - Several different approaches are considered
    - ★ VMO sends a query to VHO
    - ★ VMO queries a service that would propagate solar wind/IMF data to Earth prior to query execution
    - ★ Use and/or create propagated solar wind/IMF data products and search them directly by VMO

Activity Summary and Outlook

- The VHO/VMO Middleware prototype is available but with limited data sets
  - The VMO web interface already utilizes the latest version
  - The VHO web interface will switch to this version in summer 2007
- A development of VO Query Language has been started
  - VHO/VMO plan to employ VOQL in the API to communicate with services
  - VOQL is important to the entire VxO environment
- VHO/VMO propose to start and/or continue using the open source development approach because in their experience it provides important benefits:
  - Faster development
  - Wider community acceptance
  - More testing
- Major tasks planned for 2007:
  - Finish testing the Middleware and deploy it at both VMO and VHO sites.
  - Develop and deploy a first version of VOQL.
  - Initiate descriptions of new data products and register them with VMO or VHO.
  - Focus on optimization of the database structure and query execution.
  - Improve web interfaces to increase user comfort and acceptance.

Activity Summary and Outlook

- The VHO/VMO Middleware prototype is available but with limited data sets
  - The VMO web interface already utilizes the latest version
  - The VHO web interface will switch to this version in summer 2007
- A development of VO Query Language has been started
  - VHO/VMO plan to employ VOQL in the API to communicate with services
  - VOQL is important to the entire VxO environment
- VHO/VMO propose to start and/or continue using the open source development approach because in their experience it provides important benefits:
  - Faster development
  - Wider community acceptance
  - More testing
- Major tasks planned for 2007:
  - Finish testing the Middleware and deploy it at both VMO and VHO sites.
  - Develop and deploy a first version of VOQL.
  - Initiate descriptions of new data products and register them with VMO or VHO.
  - Focus on optimization of the database structure and query execution.
  - Improve web interfaces to increase user comfort and acceptance.

Activity Summary and Outlook

- The VHO/VMO Middleware prototype is available but with limited data sets
  - The VMO web interface already utilizes the latest version
  - The VHO web interface will switch to this version in summer 2007
- A development of VO Query Language has been started
  - VHO/VMO plan to employ VOQL in the API to communicate with services
  - VOQL is important to the entire VxO environment
- VHO/VMO propose to start and/or continue using the open source development approach because in their experience it provides important benefits:
  - Faster development
  - Wider community acceptance
  - More testing
- Major tasks planned for 2007:
  - Finish testing the Middleware and deploy it at both VMO and VHO sites.
  - Develop and deploy a first version of VOQL.
  - Initiate descriptions of new data products and register them with VMO or VHO.
  - Focus on optimization of the database structure and query execution.
  - Improve web interfaces to increase user comfort and acceptance.

Activity Summary and Outlook

- The VHO/VMO Middleware prototype is available but with limited data sets
  - The VMO web interface already utilizes the latest version
  - The VHO web interface will switch to this version in summer 2007
- A development of VO Query Language has been started
  - VHO/VMO plan to employ VOQL in the API to communicate with services
  - VOQL is important to the entire VxO environment
- VHO/VMO propose to start and/or continue using the open source development approach because in their experience it provides important benefits:
  - Faster development
  - Wider community acceptance
  - More testing
- Major tasks planned for 2007:
  - Finish testing the Middleware and deploy it at both VMO and VHO sites.
  - Develop and deploy a first version of VOQL.
  - Initiate descriptions of new data products and register them with VMO or VHO.
  - Focus on optimization of the database structure and query execution.
  - Improve web interfaces to increase user comfort and acceptance.